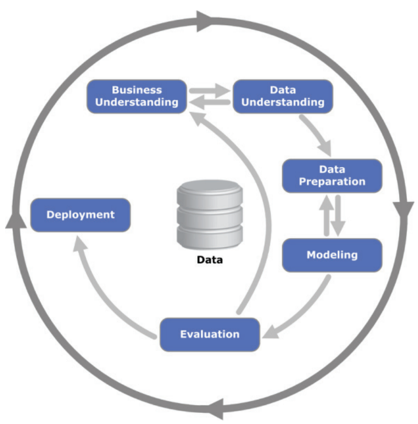
MACSEA Data Science Portfolio

Data science, machine learning, and artificial intelligence are the engines of the future. They're already at work all around us, in every aspect of our lives. MACSEA has decades of experience in delivering artificial intelligence, machine learning, and predictive analytics solutions, through neural networks, deep learning, and other data mining techniques across a wide range of industrial and military applications. The common thrust across our projects has been data-driven machine learning for creating artificial intelligence. We stand ready to help you on your journey towards Digitalization!

In working with clients, we adhere to the Cross-Industry Standard Process for Data Mining (CRISP-DM), which includes these key activities:

- Business understanding & goals
- Data understanding
- Data preparation
- Modeling & machine learning
- Model testing & evaluation
- Model deployment



Together, we will clarify your requirements, establish project goals, design a project plan for success, and assist you in capitalizing on the digitalization revolution.

Data Mining with RapidMiner

Our staff is conversant in a number of programming languages and data science tools, such a C, C++, C#, VB, JavaScript, PHP, Python, databases, Excel, and MatLab. We are a Microsoft Developer and an Ambassador Partner with RapidMiner, a premiere data mining software company.

Partner with us for your next data science project!

Initial Consult on us - We'll provide consulting time to understand your business problem, objectives, and possible solutions, followed by a formal written proposal for project engagement. You can decide to proceed with the project or not.

Project engagement – If you like our proposal, you can contract us for project engagement. We ask for a small get-started commitment, with subsequent payments tied to project milestones.

Contact Info:

Kevin Logan – President/CEO
klogan@macsea.com
860-535-3885 (office)
860-861-3172 (mobile)

This sampling of projects covers a range of real-world applications we have performed. A common thrust is data-driven machine learning for creating artificial intelligence.

Click on a picture below to jump to the project description. Click on project heading to return here.

Human Activity Recognition

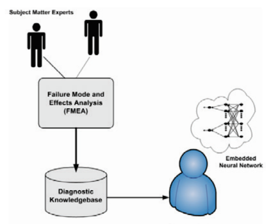
Exercise Classification from Wearable Sensors

Optimal Ship Transit Planning

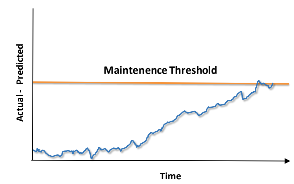
Autonomous Ship Navigation

Sensor Accommodation for Critical Control Systems

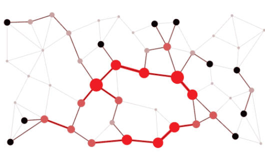
Machine Learning of Dynamic Diesel Engine Behavior



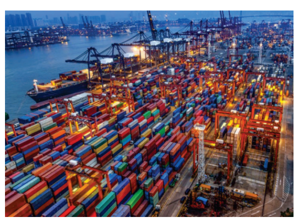
Intelligent Software Agents for Condition-Based Maintenance



Digital Twins for Equipment Health Monitoring



Communications Network Optimization

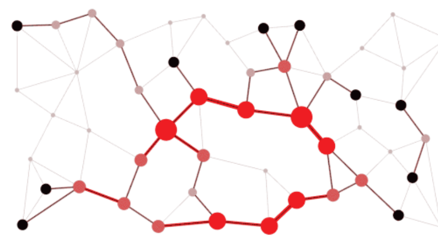


Container Routing Optimization

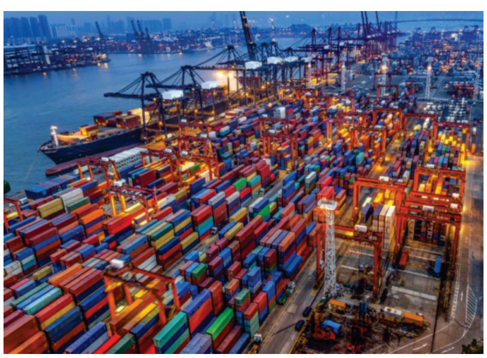
Communications Network Optimization

In order to establish commercial feasibility, NASA wanted to predict the market capture rate of its new satellite technology against competing voice and data carriers in domestic communications networks. Carrier pricing structures involved both distance-based and volume-based service costs, which increased the complexity of the modeling solution.

We developed a network analytical model that optimizes hybrid networks involving combinations of distance-sensitive and traffic-sensitive services. Least-cost routing and clustering algorithms were combined to determine minimum-cost network routing solutions. NASA was able to project traffic capture for postulated service rates necessary to provide ROI on the advanced satellite technology it was developing.



Container Routing Optimization



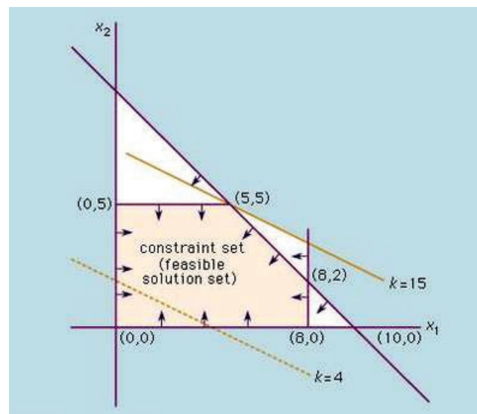
An ocean container shipper was providing "through" rate quotations to its customers, which included the combined price for the ocean, rail, and truck segments of a container move. Because of rate deregulation, the rail and truck container movement prices were frequently changing. The shipping company needed to maximize its profits when providing fixed-price, through rate transportation services to customers, taking advantage of least-cost inland routing in a dynamic pricing environment for contracted rail and truck services.

We developed a container routing optimization model generating both the least-cost and least-time routing solutions for the rail and truck transport segments. The network analysis algorithms removed the need for human analysis, reducing the complex analytical pricing process to a simplified rate data management process. The model was integrated into the carrier's container movement control system, allowing incoming customer cargo movement requests to be automatically priced, routed, and authorized using the least-cost/least-time inland rail and truck carriers.

Optimal Ship Transit Planning for Energy Conservation

Fuel costs for naval surface ships exceed a billion dollars per year. The Navy is always looking for ways to conserve fuel. Many twin-screw ships are configured with two engines on each propeller shaft. Often times, crews operate the ships with all four engines running, when the required mission speed can be achieved using fewer engines, potentially saving lots of fuel in the process.

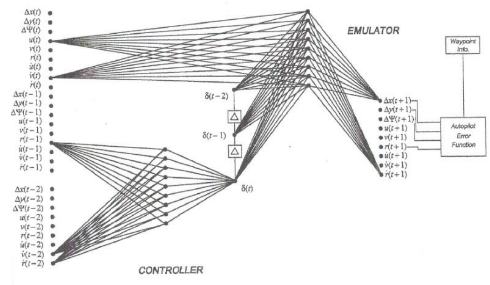
Transit planning was formulated as a classic Linear Programming problem with constraints. The objective was to determine the optimal engine configuration for a transit over a specified distance in a specified time that minimizes fuel consumption while balancing the amount of time each engine is operated (i.e. wear). Constraints included the ship's fuel storage capacity, the maximum allowed transit time, and the required transit distance. Our linear programming solution was back fitted to historical data from one ship's 14-day transit. The solution projected 208K gallons in fuel savings, worth about $700K at prevailing prices.



Autonomous Ship Navigation

Many studies have proven that most maritime accidents are due to human error. Accidents involving large ships such as oil tankers can have catastrophic environmental effects. Today's vessels are large, have massive inertia, and are slow to respond to control commands. Navigator anticipation and judgment are essential, as navigational commands must be issued well in advance of desired maneuvers. The margin for human error is very small, particularly in restricted or congested waterways. With the strong current interest in autonomous vessels, ensuring safety and reducing risks of autonomously navigating large ships is of growing concern. AI technology for autonomous ships is under development by a number of companies, but contrary to recent media stories, there's a lot more that needs to be done to ensure safety.

We developed Deep Learning neural networks to build a "neurocontroller" able to learn the maneuvering hydrodynamics of a ship and then steer it across restricted waterways. A ship simulator was used to generate neurocontroller training data, including hydrodynamic coefficients from an actual tanker. Modeled effects included hydrodynamic, propeller, and rudder forces and moments. Using simulation data, the controller was trained to accurately navigate the ship in the same manner as humans. In the simulation run shown to the right, the 864 foot-long ship's heading error from the target track was less than one degree, with a track error of less than 20 feet. If a ship sustains damage during transit, its response behavior to navigation commands may change. This technology will allow in-situ relearning of maneuvering hydrodynamics. Autonomous, self-learning ship control promises to eliminate the major cause of ship accidents: human error.

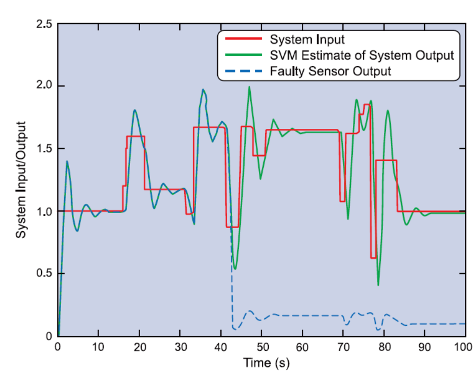


Simulation Results

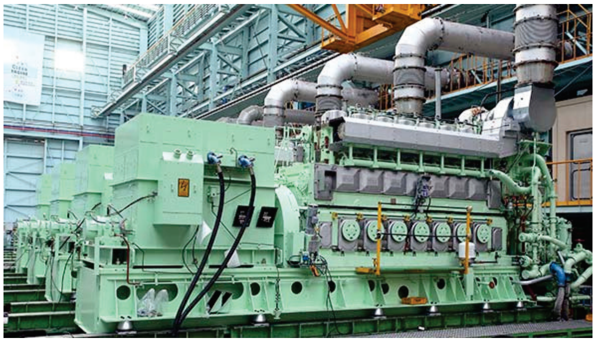
Sensor Accommodation for Military Control Systems

Sensor data quality is critical to advanced control systems on next-generation military spacecraft, aircraft, and ships. Critical functionalities, such as automatic plant reconfiguration under damage conditions, cannot be realized without accurate and reliable sensor data. Advanced algorithms and digital twins can provide "soft" sensor estimates as a substitute for lost sensors. Sensor accommodation, which includes sensor failure detection and software recovery, is a key enabling technology for all systems requiring reliable data to synthesize decisions.

We developed sensor accommodation models based on Support Vector Machine (SVM) machine learning algorithms. The model learns the nonlinear relationships between sensors of a healthy system. In the event of a sensor failure, the SVM model robustly estimates the failed signal based these learned relationships. The robustness of SVM models make them ideal for sensor analytical redundancy that will allow critical system control functions to continue operating properly under damage and failure conditions.



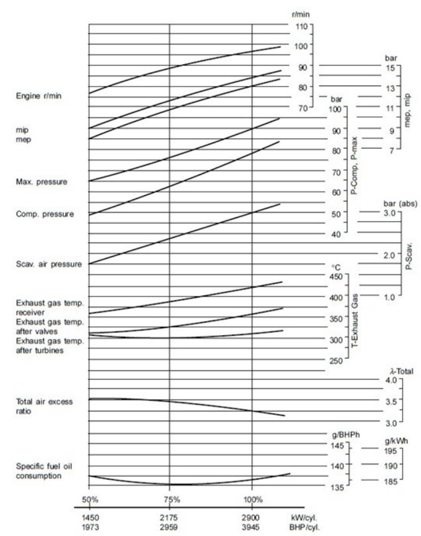
Machine Learning of Dynamic Diesel Engine Behavior

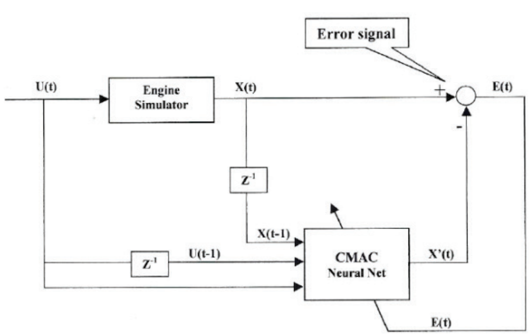


Diesels are the most widely used internal combustion engine in industry, including applications in power plants, on-road and off-road vehicles, construction and agricultural machinery, trains, and ships. The larger the engine, the more costly it is to fix when breakdowns occur and usually, the more revenue is lost during downtime. Condition-Based Maintenance (CBM) can significantly reduce down time and maintenance costs. However, this maintenance strategy relies on accurate baseline performance models to check engine condition, which are often difficult to obtain.

Engine manufacturers typically perform shop tests for each engine design and generate performance curves similar to those shown to the right. These mathematical relationships show how various performance measurements change as a function of engine power. These curves are often used as a basis of diagnosing when engine problems are starting to develop that could lead to future failures. At a given engine power level, deviations from values predicted by these curves are tell-tale symptoms of impending problems.

Shop trial curves represent performance under ideal conditions for a new engine burning high-quality fuel under controlled conditions. Actual performance in the field for a different, older, or poorly maintained engine, running on lower quality fuel and varying ambient conditions is often significantly different.





We developed AI-based models capable of continuous, real-time machine learning of diesel engine performance baselines, similar to the above curves, using a neural network known as the Cerebellar Model Articulation Controller (CMAC). The learning system was capable of “imprinting” on any engine during live operation in the field. With accurate engine performance baselines, CBM strategies can be implemented to automatically diagnose and predict failures, while preventing downtime and reducing maintenance costs.

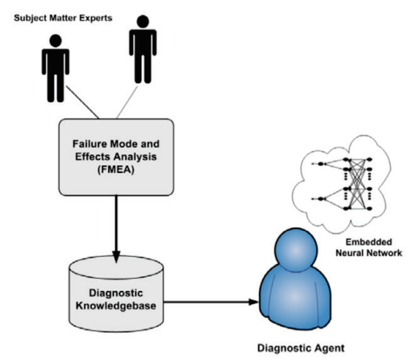
Intelligent Software Agents for Condition-Based Maintenance

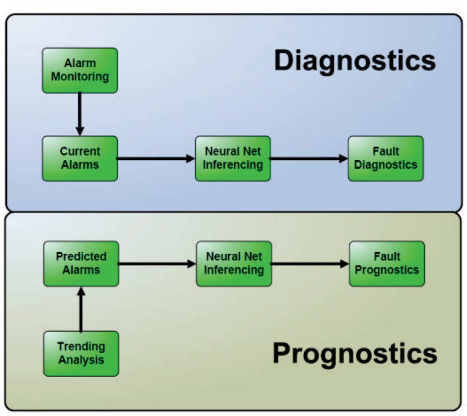
Intelligent software agents can automatically and tirelessly monitor, diagnose, and predict equipment faults from vast volumes of data.

A Condition-Based Maintenance (CBM) strategy relies on data to determine the need for equipment maintenance. Maintenance is conducted when sensor data shows degrading performance. Using CBM to drive maintenance decisions has proven to reduce costs by 30% and eliminate breakdowns by up to 70% in several industries. With the proliferation of modern automation and industrial IoT, companies will be hard-pressed to extract value from this volume of data without automating the analytical processes.

We created diagnostic and prognostic software agents to continuously monitor equipment health from real-time sensor data streams. At the heart of the system, performance baselines of the monitored physical assets are used to track anomalous equipment behavior. Salient data attributes are classified as high, low or normal and these classifications are used for real-time diagnostics. Data is also automatically trended over time to predict future alarm conditions, which are then used by prognostics software agents for equipment failure prediction.

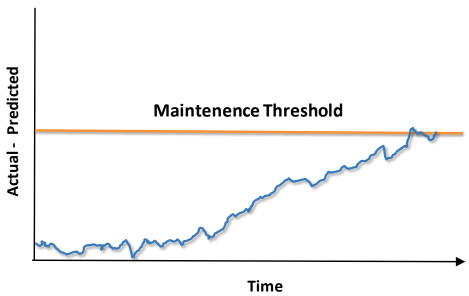
For implementation, subject matter experts perform a Failure Mode and Effects Analysis of the equipment as part of our knowledge acquisition process. The resulting fault-symptom matrix is used to train probabilistic neural network models that become the embedded intelligence of diagnostic and prognostic software agents. The fault-symptom matrix relates faults to symptom patterns, expressed as high, low, or normal performance deviation attributes. The neural network learns these pattern associations.





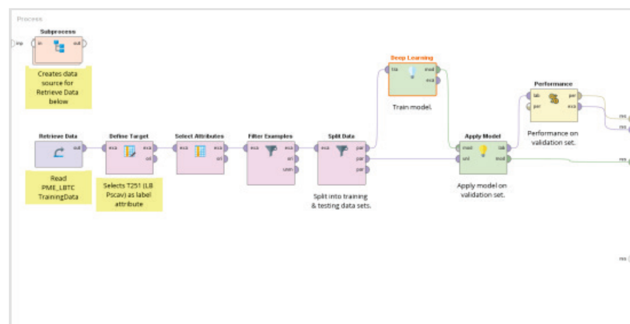
Large numbers of agents can be dispatched as needed throughout machinery plant IT infrastructures. Integration with existing automation provides the real-time input sensor data streams to the agents. The agents report equipment health status back to the automation via API calls or custom interface modules. The net result is plant-embedded artificial intelligence, which is particularly valuable for legacy, outdated automation systems. CBM can then be used to drive maintenance decisions to reduce costs and eliminate breakdowns.

Digital Twins for Equipment Health Monitoring

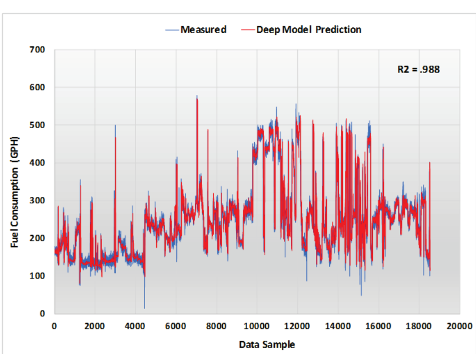


A Digital Twin is defined as a computer model of the behavior of a physical system, object, or process that helps optimize a specific aspect of business performance. We are building digital twins for predictive analytics in support of condition-based and reliability-centered maintenance strategies. Digital twins are used to model equipment behaviors and track deviations between actual and predicted performance over time. When performance deviations exceed defined maintenance thresholds, preventive or corrective maintenance actions can be performed to restore normal equipment behavior, thus avoiding failures.

We are developing Deep Learning (DL) models to create a number of digital twins from actual equipment sensor data. The RapidMiner Studio software platform facilitates rapid development and deployment of digital twin models in real world environments. Large datasets of historical equipment operating data can be quickly processed to monitor any number of physical assets. This data-driven approach can deliver many benefits, such as avoiding failures and business losses, increasing equipment reliability and availability, and reducing maintenance costs.



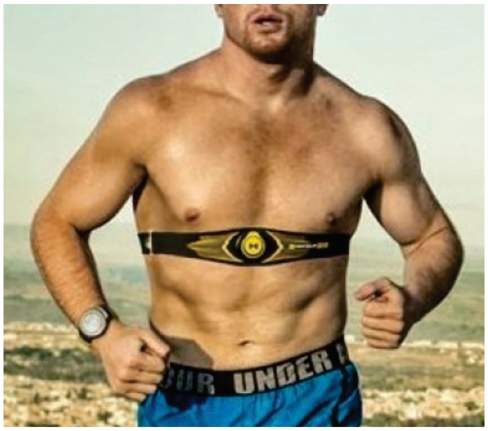
RapidMiner Process



High-accuracy Deep Model Predictions

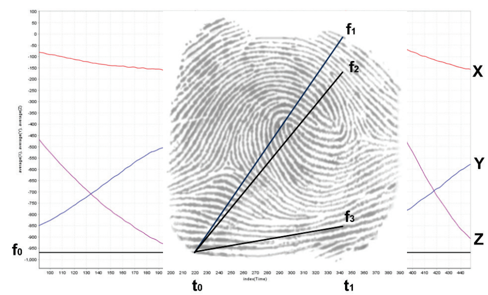
In one recent use case, we deployed a digital twin to track fuel consumption on large diesel engines typically used for propulsion on commercial and naval ships. The business goal was to detect excessive consumption as early as possible so that corrective action can be taken to restore optimal efficiency. Even small percentage savings add up to substantial money annually, particularly when aggregated over a large fleet of engines. We trained a deep learning model to build the diesel engine digital twin, regressing fuel consumption values against 85 other data attributes from 327,000 historical records. The trained model explained over 98% of the variability in the training dataset.

Exercise Classification from Wearable Sensors



If you monitor your heart rate with a wearable like Fitbit, you're probably interested in achieving a stronger heart through exercise. A major sports apparel company selling a heart and body monitoring chest strap wanted new machine learning algorithms that could identify specific exercises the athlete was performing. Heart rate could then be correlated to different exercises and more meaningful training feedback could be provided for faster health improvement.

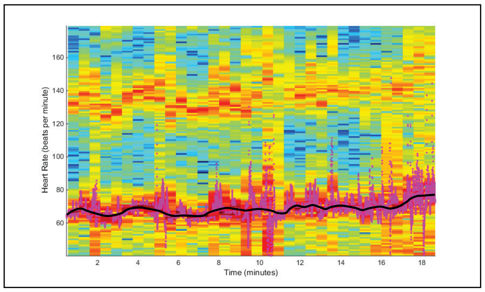
Using accelerometer data from the chest strap, we engineered a novel feature set that exploited patterns in the exercise data. We call this an "ExcerPrint' (as in fingerprint), because the new attributes were similar to a fingerprint, forming a unique pattern for each different exercise.



ExcerPrint Feature Set

At the gym, we used the chest strap to generate data from seven different exercises and then trained classifiers to identify them. Eleven machine learning models were developed, including various neural networks, decision trees, random forests, naïve Bayes, and nearest neighbor techniques. The models were tested against unlabeled gym exercise data. Average predictive accuracy was 96% across the various models. The resulting machine learning classifier was able to imprint on each individual athlete very successfully.

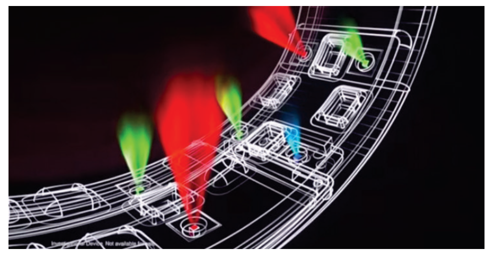
Human Activity Recognition



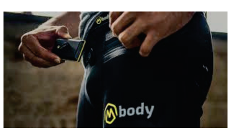
Smartphones, wearables, and ambient intelligence driven by IoT and in-home connected devices such as Alexa provide Petra-bytes data. New Apps are transforming the way we live. There's plenty of opportunity on the horizon for further disruptive innovation. We are pivoting our AI-based equipment health monitoring technology into human activity monitoring. We see this as a huge growth area and are actively performing R&D in a few select areas.

Digital Health

We are using data from optical PPG sensors integrated into most Fitbit-type wearables to develop AI and advanced signal processing algorithms to extract heart rate, respiration rate, and hydration-related biomarkers. We're not market ready yet, but key target markets are very large and include the military, first responders, such as firefighters, athletes, infants, the elderly, and the chronically ill.



Smart Clothing



Conductive fabrics and sensorized clothing technologies have been growing fast, with the involvement of nearly every major sports apparel company. We have been developing stress and emotion sensing technology using wearable sensor data, including heart rate, respiration rate, heart rate variability, skin temperature, galvanic skin response, and motion. Machine learning allows self-ranking models to be trained on each individual.

Sports Performance Training

Wearable sensor technology is becoming in widespread use across almost all sports for training and performance improvement. We developed a wearable training solution for bowling that includes a sensorized glove and shot analyzer App to help bowlers perform more consistently and raise their bowling average. The patent-pending analyzer uses artificial intelligence to automatically detect a shot, compute key shot metrics from sensor data, and analyze bowler motion data.